Reinforcement Learning

Nipun Batra

IIT Gandhinagar

July 30, 2025

Forecasting

Nipun Batra

July 19, 2025

IIT Gandhinagar

Acknowledgement

Forecasting: Principles and Practice

Some famous forecasts

- I think there is a world market for maybe five computers. (Chairman of IBM, 1943)

Some famous forecasts

- I think there is a world market for maybe five computers. (Chairman of IBM, 1943)
- Computers in the future may weigh no more than 1.5 tons. (Popular Mechanics, 1949)

Some famous forecasts

- I think there is a world market for maybe five computers. (Chairman of IBM, 1943)
- Computers in the future may weigh no more than 1.5 tons. (Popular Mechanics, 1949)
- There is no reason anyone would want a computer in their home. (President, DEC, 1977)

Some famous forecasts

- I think there is a world market for maybe five computers. (Chairman of IBM, 1943)
- Computers in the future may weigh no more than 1.5 tons. (Popular Mechanics, 1949)
- There is no reason anyone would want a computer in their home. (President, DEC, 1977)
- New open standards created in the mobile era, such as HTML5, will win on mobile devices (and PCs too). Perhaps Adobe should focus more on creating great HTML5 tools for the future, and less on criticizing Apple for leaving the past behind. (Steve Jobs about Flash in 2010)

Applications

- Whether to build a power plant in the next five years based on forecast of future demand

Applications

- Whether to build a power plant in the next five years based on forecast of future demand
- Stocking an inventory based on forecast of stock requirements

3 Factors affecting predictability of an event

- how well we understand the factors that contribute to it;

3 Factors affecting predictability of an event

- how well we understand the factors that contribute to it;
- how much data is available;

3 Factors affecting predictability of an event

- how well we understand the factors that contribute to it;
- how much data is available;
- whether the forecasts can affect the thing we are trying to forecast.

3 Factors affecting predictability of an event

For electricity demand forecasting:

- we have a good undertsanding that the demand is largely a function of the temperature

3 Factors affecting predictability of an event

For electricity demand forecasting:

- we have a good undertsanding that the demand is largely a function of the temperature
- there is a lot of past data available

3 Factors affecting predictability of an event

For electricity demand forecasting:

- we have a good undertsanding that the demand is largely a function of the temperature
- there is a lot of past data available
- the fact that we can forecast electricity demand does not seem to affect the forecast

3 Factors affecting predictability of an event

For stock price prediction:

- we do not have a good undertsanding of the underlying process

3 Factors affecting predictability of an event

For stock price prediction:

- we do not have a good undertsanding of the underlying process
- there is a lot of past data available

3 Factors affecting predictability of an event

For stock price prediction:

- we do not have a good undertsanding of the underlying process
- there is a lot of past data available
- the fact that we can forecast stock prices will lead to change in market dynamics and it will affect the forecast

Example of Forecasting

Forecast of production of beer in Australia

Dark blue lines show the mean forecast
Light blue band shows the confidence interval

Three Types of Forecasting Models

Task: Forecast Electricity Demand at Time T

1. Explanatory model
   $E_T =$
   $f(Temperature_T, GDP_T, Population_T, Day_T, Month_T, Hour_T)$

Three Types of Forecasting Models

Task: Forecast Electricity Demand at Time T

1. Explanatory model
   $E_T =$
   $f(Temperature_T, GDP_T, Population_T, Day_T, Month_T, Hour_T)$
2. Timeseries model
   $E_T = f(E_{T-1}, E_{T-2}, \cdots)$

Three Types of Forecasting Models

Task: Forecast Electricity Demand at Time T

1. Explanatory model
   $E_T =$
   $f(Temperature_T, GDP_T, Population_T, Day_T, Month_T, Hour_T)$
2. Timeseries model
   $E_T = f(E_{T-1}, E_{T-2}, \cdots)$
3. Mixed model
   $E_T = f(E_{T-1}, E_{T-2}, Temperature_T, GDP_T, Population_T, \cdots)$

The Importance of Domain Knowledge in Forecasting

Weekly economy passenger load on Ansett Airlines.

- There was a period in 1989 when no passengers were carried

The Importance of Domain Knowledge in Forecasting

Weekly economy passenger load on Ansett Airlines.

- There was a period in 1989 when no passengers were carried
- There was a period of reduced load in 1992. This was due to a trial in which some economy class seats were replaced by business class seats.

The Importance of Domain Knowledge in Forecasting

Weekly economy passenger load on Ansett Airlines.

- There was a period in 1989 when no passengers were carried
- There was a period of reduced load in 1992. This was due to a trial in which some economy class seats were replaced by business class seats.
- A large increase in passenger load occurred in the second half of 1991.

The Importance of Domain Knowledge in Forecasting

Weekly economy passenger load on Ansett Airlines.

- There was a period in 1989 when no passengers were carried
- There was a period of reduced load in 1992. This was due to a trial in which some economy class seats were replaced by business class seats.
- A large increase in passenger load occurred in the second half of 1991.
- There are some large dips in load around the start of each year. These are due to holiday effects.

Time Series Patterns

- **Trend**: Long term increase or decrease.

Time Series Patterns

- **Trend**: Long term increase or decrease.
- **Season**: Time series is affected by seasonal factors such as the time of the year or the day of the week

Time Series Patterns

- **Trend**: Long term increase or decrease.
- **Season**: Time series is affected by seasonal factors such as the time of the year or the day of the week
- **Cyclic**: A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency.

Seasonal v/s Cyclic

- Seasonality: within every year, there is pattern of sales

Seasonal v/s Cyclic

- Seasonality: within every year, there is pattern of sales
- Cyclic: Every 6 years or so, there is a similar pattern

No Pattern in Timeseries

- Trend: None exists

Thus, very difficult to forecast

No Pattern in Timeseries

- Trend: None exists
- Seasonality: None exists

Thus, very difficult to forecast

No Pattern in Timeseries

- Trend: None exists
- Seasonality: None exists
- Cyclic: None exists

Thus, very difficult to forecast

Simple Forecasting Methods

- Average method: $\hat{y}_{T+h}|y_{1:T} = \frac{\sum_{t=1}^{T} y_t}{T}$

Simple Forecasting Methods

- Average method: $\hat{y}_{T+h}|y_{1:T} = \frac{\sum_{t=1}^{T} y_t}{T}$
- Naive method: $\hat{y}_{T+h}|y_{1:T} = y_T$

Simple Forecasting Methods

- Average method: $\hat{y}_{T+h}|y_{1:T} = \frac{\sum_{t=1}^{T} y_t}{T}$
- Naive method: $\hat{y}_{T+h}|y_{1:T} = y_T$
- Seasonal naive method:

Simple Forecasting Methods

- Average method: $\hat{y}_{T+h}|y_{1:T} = \frac{\sum_{t=1}^{T} y_t}{T}$
- Naive method: $\hat{y}_{T+h}|y_{1:T} = y_T$
- Seasonal naive method:
    - Same as naive method, but, incorporates seasonal information, e.g. Forecast is the same as the value in the same month last year.

Simple Forecasting Methods

- Average method: $\hat{y}_{T+h}|y_{1:T} = \frac{\sum_{t=1}^{T} y_t}{T}$
- Naive method: $\hat{y}_{T+h}|y_{1:T} = y_T$
- Seasonal naive method:
  - Same as naive method, but, incorporates seasonal information, e.g. Forecast is the same as the value in the same month last year.
  - $\hat{y}_{T+h}|y_{1:T} = y_{T+h-m(k+1)}$ where m is the seasonal period and k is is the integer part of $\frac{(h-1)}{m}$

Simple Forecasting Methods

- Average method: $\hat{y}_{T+h}|y_{1:T} = \frac{\sum_{t=1}^{T} y_t}{T}$
- Naive method: $\hat{y}_{T+h}|y_{1:T} = y_T$
- Seasonal naive method:
  - Same as naive method, but, incorporates seasonal information, e.g. Forecast is the same as the value in the same month last year.
  - $\hat{y}_{T+h}|y_{1:T} = y_{T+h-m(k+1)}$ where m is the seasonal period and k is is the integer part of $\frac{(h-1)}{m}$
  - Let us assume we are forecasting monthly and we want to forecast for Mar 2020 (month=3) and current time is Jan 2020 (month=1). Thus, $h = 2$. Let us assume yearly seasonality, i.e. $m = 12$. Thus, prediction for Mar 2020 is value at (Jan+2 months) - $12 \times 2 - 1\%12$ =Mar 2020 - 12 Months = Mar 2019

Simple Forecasting Methods

Learning: Simple solutions often work well, especially if you know about the domain.

Evaluating Forecast Accuracy

Timeseries cross-validation for 1 timestep ahead prediction

Question: How do you nested CV?
Answer: Similarly divide the train into train and validation preserving the notion of timeseries.

Evaluating Forecast Accuracy

Timeseries cross-validation for $k = 4$ timestep ahead prediction

Stationarity

- A stationary time series is one whose properties do not depend on the time at which the series is observed

Stationarity

- A stationary time series is one whose properties do not depend on the time at which the series is observed
- Time series with trends, or with seasonality, are not stationary

Stationarity

- A stationary time series is one whose properties do not depend on the time at which the series is observed
- Time series with trends, or with seasonality, are not stationary
- White noise series is stationary — it does not matter when you observe it, it should look much the same at any point in time.

Stationarity: Which of these are stationary?

- Series with trends: a, c, e, f, i

Stationarity: Which of these are stationary?

- Series with trends: a, c, e, f, i
- Series with seasonality: d, h, i

Stationarity: Which of these are stationary?

- Series with trends: a, c, e, f, i
- Series with seasonality: d, h, i
- Stationary: b and g (cycles are aperiodic)

Differencing

What is the relation between a and b?

b is the first order time difference of a!

b is stationary, while a is not!

Differencing

For (a) the ACF is significant
For (b), the ACF declines rapidly

Autoregression

$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$,where ε_t is white noise.